

Original research article

Exploring constraints on the realised value of a forecast-based climate service

Edward C.D. Pope^{a,*}, Carlo Buontempo^{a,b}, Theo Economou^{a,c}^a Met Office, Hadley Centre, Fitzroy Road, Exeter, Devon EX1 3PB, UK^b ECMWF, Shinfield Park, Reading, UK^c Department of Mathematics, University of Exeter, New North Road, Exeter EX4 4QE, UK

A B S T R A C T

The increasingly widespread use of climate services for decision-making has highlighted the need for service developers to more clearly establish the benefits and limitations of the information they provide. Using a simple cost/loss framework applied to an idealised forecast system, we explore the critical level of accuracy required for the expected utility of a predictive service to exceed a benchmark based on the climatological frequency of the hazard. In this simplest case, the critical accuracy is a function only of the cost/loss ratio for the decision and the climatological frequency. Although the utility of climate services depends on a number of non-climate-related factors, comparing current forecasting capabilities to an estimate of the critical accuracy can provide a guide to the expected marginal benefit of the climate service. More generally, to ensure that climate service evaluation is relevant to users it must account for the finite duration over which a climate service is used, since the realised performance can differ significantly from its expectation. Here, the statistical distribution of realised performance is used to quantify the prior likelihood that the service will outperform a benchmark as a function of accuracy, climatological frequency, cost/loss ratio and duration of usage. This information can help users assess the potential risks associated with adopting a climate service for a given period, and may help identify geographical regions and meteorological parameters for which a forecasting tool could provide a worthwhile investment with respect to the next-best alternative.

Practical Implications.

Demonstrating the performance of seasonal forecasts in terms of user-relevant metrics is essential for critically assessing the utility of climate services for real-world applications. Doing so should both improve user confidence in the service and highlight where improvements in forecasting capabilities could provide tangible societal benefit. Using a decision-theoretic framework, we have analysed the characteristics required for a deterministic binary forecast service to outperform a standard benchmark. In this case the benchmark is based on knowing only the climatological frequency of the weather hazard, which provides a simple and unequivocal reference. The general approach demonstrated here can also be applied for any other benchmark forecast.

Rather than focusing on the expected (i.e. long-run average) economic value of the system, this analysis explores the distribution of economic value that could be realised given that a real seasonal forecast system is only used for a limited duration, e.g. once annually for 5 years. For the forecast system modelled here, the shape of the distribution is governed by cost/loss ratio for the decision, forecast accuracy, climatological frequency, the duration over which forecasts are used, and whether the user always follows the forecast guidance. This distribution forms a basis for understanding the prior likelihood that, over a given test

period, a seasonal forecasting service will provide user benefits beyond those expected from current approaches to decision-making. Where possible, estimating the likelihood of additional benefit would provide important context for user experience, and demonstrate that service providers understand the implications of using their service. In turn, this may facilitate more profound engagement with users, leading to improved development of a usage of climate services.

1. Introduction

Over recent years there has been an increasing effort to translate climate science into services, with the aim of providing information that has the potential to improve decision-making across a range of time-scales. Notable examples include the Global Framework for Climate Services (e.g. Hewitt et al., 2012), the sectoral information systems developed within Copernicus Climate Change Services for the energy and water sectors, as well as a number of H2020 projects such as S2S4E (energy), MEDGOLD (Agriculture) or IMPREX (water). Given the recent emphasis on service development, and the potential for widespread use (e.g. Thornes and Stephenson, 2001; Meza et al., 2007; Goddard et al., 2010; Palin et al., 2016; Bett et al., 2017; Henley and Pope, 2017;

* Corresponding author.

E-mail address: edward.pope@metoffice.gov.uk (E.C.D. Pope).<https://doi.org/10.1016/j.cliser.2019.100102>

Received 20 August 2018; Received in revised form 8 April 2019

2405-8807/ © 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Vaughan et al., 2018), it is vital to ascertain the potential benefits and limitations of the information being generated. At the simplest level, this requires an assessment of the relationship between forecast performance and value (however this is defined) to a user (e.g. Brooks, 2013). Critically, this relationship depends on how the information is used (e.g. Bruno Soares and Dessai, 2015; Buontempo et al., 2017; Golding et al., 2017; Hewitt et al., 2017), meaning that improvements in forecast accuracy need not automatically lead to a more valuable service. For this reason, there is a clear need for service developers to understand the impact to users of forecast errors (e.g. Mason, 2004), the duration over which the service will be used, and the typical cost/loss ratios associated with the decisions.

Achieving a truly objective measure of service benefit is a complex challenge for a variety of reasons. Firstly, it is often difficult to quantify the costs and losses (direct or indirect) of a specific decision. Secondly, while many decisions are binary (e.g. act or do not act), others are more nuanced, requiring differing degrees of action (e.g. Falloon et al., 2018). In terms of a cost/loss framework (e.g. Lindley, 1985), the cost of making a specific decision (e.g. to grit the roads) may be relatively fixed. However, the losses associated with incorrect action can be highly subjective, incorporating both financial and reputational losses. As an example, more risk-averse decision-makers tend to subjectively assign greater losses due to inaction than less risk-averse individuals (e.g. Blench, 1999). Accordingly, the typical cost/loss ratio for a single decision is dependent on both individual and organisational risk appetite. For this reason, it is probably not possible to develop a single approach which quantifies the performance of all user-relevant aspects of a service.

Our previous work (Pope et al., 2017) made use of a synthetic, deterministic binary forecast combined with a standard 2-by-2 cost/loss decision theoretic framework to explore a range of factors affecting the expected value of forecast information. This implicitly assumed that the work of converting a probabilistic forecast into a binary forecast had already been done. In reality, this is not a trivial task and depends on the user's risk appetite, cost/loss ratio and interpretation of the information.

To maintain tractability, we adopt the same general approach here and explore two complementary user-relevant extensions: firstly, the parameter dependence of the critical forecast accuracy required for the expected expense of a forecast-based service to be lower than that of a benchmark (which is assumed here to be the use of climatological frequencies). Secondly, the "expense" is typically described in terms of the long-run or expectation value; this can be misleading since the timescale required for the realised economic value to converge to this asymptote can be much greater than the duration over which the climate service is actually used (see Pope et al., 2017). For example, consider a deterministic binary seasonal forecast which is issued once per year and correct typically 80% of occasions. Over a 5-year trial period, we would expect the forecast to be correct four times and incorrect once. There are also non-zero probabilities that the forecast is correct or incorrect on all 5 occasions. For this reason, we use *stochastic simulations to explore the factors that affect the probability distribution of Mean Realised Value (MRV) for a plausible forecast system used over a finite period*. The MRV refers to the mean realised economic value extracted from the forecast system used for a finite time interval, e.g. a seasonal forecast used once annually for 5, 10 or 20 years. From the distribution of the MRV, we can estimate the prior likelihood that the climate service, used over an N-year interval, will outperform the benchmark. As part of this, we also explore the consequences of two decision-making strategies: i) always following the forecast guidance, regardless of the specified forecast accuracy; ii) following the forecast guidance only if the forecast accuracy exceeds a critical threshold, below which decisions are based on climatological information. For users, this prior likelihood could help provide the information necessary to understand whether a climate service is suitable for their requirements, given their risk appetite. For service developers, this

information could help to identify geographical regions and meteorological parameters for which the forecasting tool is likely to provide value to a user.

For the purposes of this investigation, the users under consideration are assumed to be individuals or organisations that make decisions which are informed by climate services. For this reason, the terms "user" and "decision-maker" are used interchangeably throughout this work.

The remainder of the paper is structured as follows: Section 2 describes the methods used to simulate a synthetic climate service. Section 3 describes the main results, and Section 4 outlines the conclusions.

2. Methods

The modelling framework described below is designed to explore the statistical distribution of Mean Realised Value (MRV), and consists of two main parts: firstly, a decision theoretic approach which encapsulates the main factors influencing user decision-making, and is based on a cost/loss matrix that quantifies consequences for each of the four possible decision-outcome combinations; secondly, stochastic simulations of the decision-outcome combinations, based on the statistical properties of an idealised forecast system. For simplicity and transparency, this approach focuses on the properties of a binary, deterministic forecast system¹.

2.1. Decision theory

The fundamental decision-theoretic approach adopted here is the same as in Pope et al (2017) which makes use of a 2-by-2 cost/loss matrix, see also Katz and Murphy (1997), Wilks (2001), Mylne (2002), Palmer (2002), Richardson (2000, 2012). In this approach there are only two outcomes: a specific uncertain weather event Y_1 will occur with probability, $p = \Pr(Y_1)$, or it will not occur, Y_2 , with probability $1-p = \Pr(Y_2)$. Similarly, there are only two decisions: to either take action before the event at some cost, or not with associated losses if the event occurs. Table 1 shows the cost/loss matrix. We note that this represents a significant simplification and often a number of hedging techniques may be available to the decision-maker. However, to maintain clarity, we are not considering those here.

As shown in Table 1, there is a financial cost, C , associated with taking action regardless of whether adverse weather conditions (Y_1) occur. In contrast, the consequences of not taking action are dependent on the weather conditions, with a loss, L , in the event of adverse conditions, and no loss otherwise. In the absence of forecast information, knowing the cost/loss ratio and climatological probability of adverse weather conditions, combined with decision theory (Lindley, 1985) provides a way to identify which single decision-making strategy minimises losses when used repeatedly over an infinitely long period. In short, the optimal decision to make is the one that minimises expected loss, i.e. pick the smaller of $pC + (1-p)C = C$ and $pL + 0 = pL$. In contrast, for a perfect forecast, the decision-maker can adopt a mixed strategy by always picking the best decision based on the forecast information (which is always correct), with expected expense of pC .

However, real forecasts are imperfect and represent "partial" information. Indeed, the long-run performance of an imperfect forecast

¹ We note that most forecasts are probabilistic rather than binary, and that this can add an extra layer of complexity in interpreting the forecasts. In particular, transforming estimated probabilities into binary decisions will depend on the risk profile of the user. Furthermore, given the lack of reliability of many prediction systems, this conversion represents a considerable challenge. However, the cost/loss approach allows users to choose a probability threshold to adjust the relative number of hits and false alarms to match their cost/loss ratio.

Table 1

Cost-loss matrix for problem with two decisions based upon an uncertain binary event.

	Take action	Don't take action
Y_1 (event occurs)	$U_{11} = C$	$U_{21} = L$
Y_2 (event doesn't occur)	$U_{12} = C$	$U_{22} = 0$

can be worse than relying on climatological information. For this reason, it makes sense to select the decision or decisions that minimise the Expected Expense for Partial Information ($EE_{Forecast}$), as shown in Eq. (2). This implies the existence of a critical accuracy threshold, above which the forecast system provides guidance that is more valuable than climatological information - these thresholds are investigated in the results section. In this case, $EE_{Forecast}$ falls between two well-defined limits (Lindley, 1985), and is dependent on forecast accuracy. The limits are:

- 1) upper – the Expected Expense for Climatological Information ($EE_{CI} = \min(C, pL)$)
- 2) lower – the Expected Expense for Perfect Information ($EE_{PI} = pC$)

To maintain generality, the economic value associated with $EE_{Forecast}$ is expressed in relative units, with respect to both perfect and climatological information. The scale ranges between 0 and 1, with 0 indicating forecast performance is no better than climatological information, and 1 indicating perfect forecast information. Following Richardson (2000), the expected economic value (V) is given by:

$$V = (EE_{CI} - EE_{Forecast}) / (EE_{CI} - EE_{PI}) \quad (1)$$

This scaling explicitly uses the climatological frequency as a reference. The same approach could be repeated for a more stringent benchmark that, for example, accounts for the state dependence of predictability. Examples include statistical modelling approaches such as Markov chains and regression models for near-term climate prediction that can outperform static climatologies without significantly increasing the computational expense of the methodology (e.g. Van den Dool, 2006; Suckling et al 2016).

For a 2-by-2 cost/loss matrix $EE_{Forecast}$ is defined as (e.g. Lindley, 1985)

$$EE_{Forecast} = \sum_{k=1}^2 Pr(X_k) \min_D \sum_{j=1}^2 U_{Dj} Pr(Y_j | X_k) \quad (2)$$

where U_{Dj} indicate the costs or losses associated with different decision-outcome combinations, while the inner sum reflects the fact that the decision to take action is based on $Pr(Y_j | X_k)$, and the outer sum is the expectation over the possible forecasts, X_k . The D subscript in Eq. (2) represents the choice between taking action or not (see Table 1). The subscripts k and j denote the forecast and event categories, with k = 1 if the event is forecast to occur (k = 2 if not forecast), and j = 1 if the event occurs (j = 2 if the event does not occur). Selecting the decisions that minimise the expected expenses ensures that $pC \leq EE_{Forecast} \leq \min(C, pL)$. This can only be achieved robustly if the user knows when the forecast system is accurate enough to follow.

Given the deterministic binary forecast considered here, X_1 denotes the event being forecast to occur, with X_2 being the opposite. Each individual forecast will be either correct (i.e. hit or correct rejection) with corresponding probabilities: $Pr(Y_1 | X_1)$ and $Pr(Y_2 | X_2)$; or incorrect (i.e. false alarm or miss) with respective probabilities: $Pr(Y_2 | X_1) = 1 - Pr(Y_1 | X_1)$ and $Pr(Y_1 | X_2) = 1 - Pr(Y_2 | X_2)$. The probabilities $Pr(Y_1 | X_1)$ and $Pr(Y_2 | X_2)$ quantify the likelihood that the predicted outcome occurs for each category. In this case, k = j = 1 corresponds to the event being forecast and occurring, while k = j = 2 corresponds to the event not being forecast and not occurring. Henceforth, these probabilities are called the forecast accuracies.

Expanding Eq. (2) gives

$$EE_{Forecast}$$

$$= Pr(X_1) \min \{ [Pr(Y_1 | X_1) U_{11} + Pr(Y_2 | X_1) U_{21}], [Pr(Y_1 | X_1) U_{21} + Pr(Y_2 | X_1) U_{22}] \} + [1 - Pr(X_1)] \min \{ [Pr(Y_1 | X_2) U_{11} + Pr(Y_2 | X_2) U_{21}], [Pr(Y_1 | X_2) U_{21} + Pr(Y_2 | X_2) U_{22}] \} \quad (3)$$

where $U_{11} = U_{12} = C$, $U_{21} = L$ and $U_{22} = 0$ in Table 1.

For clarity of exposition, we assume that the forecast is well calibrated, meaning that the frequency bias is unity, i.e. $Pr(X_k) = Pr(Y_k)$ for k = 1, 2. Using $Pr(X_1) = Pr(Y_1) = p$, Eq. (3) then reduces to

$$EE_{Forecast} = (p) \min \{ C, L \times Pr(Y_1 | X_1) \} + (1 - p) \min \{ C, L [1 - Pr(Y_2 | X_2)] \} \quad (4)$$

The frequency bias assumption constrains the values of $Pr(Y_1 | X_1)$ and $Pr(Y_2 | X_2)$. Specifically, the definition of joint probabilities requires $\sum_{k=1}^2 Pr(Y_j & X_k) = \sum_{k=1}^2 Pr(Y_j | X_k) Pr(X_k) = Pr(Y_j)$; thus, for j = 1, $Pr(Y_1) = p = p Pr(Y_1 | X_1) + (1 - p) [1 - Pr(Y_2 | X_2)]$. Rearranging this shows that the forecast accuracies are related by the following expression

$$Pr(Y_2 | X_2) = \{1 - p [2 - Pr(Y_1 | X_1)]\} / (1 - p) \quad (5)$$

In turn, Eq. (5) demonstrates that, for a frequency bias of unity, the accuracies are identical only for $p = 1/2$. Similarly, $Pr(Y_1 | X_1) = 1$ requires that $Pr(Y_2 | X_2) = 1$ regardless of p. We note that, if $Pr(Y_1 | X_1) = 0$, $Pr(Y_2 | X_2) \neq 0$ unless $p = 1/2$.

A further implication of Eq. (5) is that $Pr(Y_2 | X_2) < 0$ for $p > 1/[2 - Pr(Y_1 | X_1)]$; thus, the model is only valid for climatological probabilities in the range $0 < p < 1/[2 - Pr(Y_1 | X_1)]$. For this reason, $Pr(Y_1 | X_1)$ is not always a good measure of accuracy. Nevertheless, the model remains valid for most applications; for example, as $Pr(Y_1 | X_1) \rightarrow 1$, the maximum allowed climatological probability tends to 1.

As shown in the Appendix, for a forecast system where $Pr(Y_1 | X_1) = Pr(Y_2 | X_2)$ the minimum value of $Pr(Y_1 | X_1)$ by construction is $1/2$, and the forecast properties are mathematically symmetrical around $Pr(Y_1 | X_1) = 1/2$. However, this is not true when $Pr(Y_1 | X_1) \neq Pr(Y_2 | X_2)$, for which $Pr(Y_1 | X_1) \rightarrow 1$ gives different forecast performance than $Pr(Y_1 | X_1) \rightarrow 0$.

More generally, Eq. (5) demonstrates that the forecast accuracies for this system can be described completely in terms of $Pr(Y_1 | X_1)$ and p, without the need for an additional free parameter.

2.2. Simulating forecasts

Over a long time-scale, the mean expense of decisions based on forecast advice will asymptotically tend towards $EE_{Forecast}$. However, over a finite interval, the mean expense can differ significantly, as determined by forecast performance event by event. One of the main aims of this paper is to explore this effect using stochastic simulations. For example, consider a forecast system that is known to correctly predict the occurrence of a binary event 4 times out of 5, on average. Calculating the prior probability of the four possible decision-outcome combinations requires knowledge of the event's climatological probability and either the forecast accuracy for non-events, or the frequency bias. Assuming the forecast accuracy for category 1 is $Pr(Y_1 | X_1) = 0.8$ and that the climatological probability is $Pr(Y_1) = p = 0.3$, the prior probability of a "hit" is $Pr(Y_1 \& X_1) = Pr(Y_1 | X_1) Pr(X_1) = 0.8 \times 0.3 = 0.24$. According to Eq. (5), and assuming frequency bias of unity, the forecast accuracy for category 2 must be $Pr(Y_2 | X_2) \approx 0.914$, meaning that the prior probability of a "miss" is $Pr(Y_1 \& X_2) = Pr(Y_1 | X_2) Pr(X_2) = (1 - 0.914) \times (1 - 0.3) = 0.06$; the prior probability of a "false alarm" is $Pr(Y_2 \& X_1) = Pr(Y_2 | X_1) Pr(X_1) = (1 - 0.8) \times 0.3 = 0.06$, and the prior probability of a "correct rejection" is $Pr(Y_2 \& X_2) = Pr(Y_2 | X_2) Pr(X_2) = 0.914 \times (1 - 0.3) = 0.64$. This situation can be modelled efficiently by simulating a time series of forecast-outcome combinations. The combinations are generated by randomly sampling from

a 4-category multinomial distribution with the appropriate probabilities for hits, misses, false alarms and correct rejections. The procedure is repeated to generate 10,000 realisations of forecast sequences. For each realisation and forecast sequence length, we calculate the Mean Realised Value (MRV) as a function of the cost/loss ratio, forecast accuracy and climatological frequency.

To provide a fair benchmark, we also simulate the corresponding sequence of forecast-outcome combinations assuming decision-making is based on climatological information only. As an example, if the cost/loss ratio is less than the climatological probability (i.e. $C/L < p$), the rational choice for decisions based only on the climatological probability is to always take action, with cost, C . Thus, the realised expense is always equal to the expected expense. However, if the cost/loss ratio exceeds the climatological probability (i.e. $C/L > p$), the rational choice for decisions based only on the climatological probability is to always not take action, with loss, L , if the event occurs, and 0 if it does not. In this case, the realised expense over a finite sequence of events can differ significantly from the expected expense, pL .

To complete the model, we consider two types of decision-maker who implement the forecast guidance in different ways:

1. Type 1 - users who always follow the forecast guidance regardless of the forecast system's specified accuracy.
2. Type 2 - users who follow the forecast advice only if the specified forecast accuracy is above a critical threshold, defined in the next section, and who otherwise base their decisions on climatological information.

In this modelling approach, users of Type 1 take action whenever simulated hits and false alarms occur, since the forecast predicted the adverse weather event to occur. For simulated misses and correct rejections, we assume that Type 1 users do not act, since the forecast predicted that the event would not occur.

Users of Type 2 are assumed to follow decision-making based on climatological information, unless the forecast accuracy exceeds the critical threshold defined in the next section. If the forecast accuracy exceeds the critical threshold, decision-making follows the same rules as for Type 1 users. While users may not always behave exactly as either Type 1 or 2, the effect of user choice in deciding when to follow forecast advice is likely to be dependent on forecast skill and their own subjective judgement. The results presented here show that the appropriate use of forecast information can limit the incidence of significant losses.

2.3. Critical accuracy

As shown in Eqs. (2) and (3), for a given cost/loss ratio and climatological probability, $EE_{Forecast}$ is a function of forecast accuracy. In particular, there is an implicit critical accuracy below which the forecast offers no expected benefits relative to the climatological benchmark. Above the critical accuracy, the expected economic value of the forecast relative to the benchmark is linearly dependent on the accuracy. Based on the mathematical definition of the $EE_{Forecast}$ (see Eqs. (2) and (3)), using the climatological frequency as the benchmark, the critical accuracy can be expressed in terms of the cost/loss ratio (C/L) and the climatological frequency (p). For brevity in the following equations, $Pr(Y_1|X_1)$ is written as α_1 , and $Pr(Y_2|X_2)$ as α_2 .

The first bracket in Eq. (3) represents the choice between action and inaction, given that the event is forecast to occur. For $\alpha_1 > C/L$, it follows that $C < L\alpha_1$ and the rational decision is to take action. As such, the critical accuracy for this case is

$$\alpha_1^* = C/L \quad (6)$$

A forecast accuracy above (below) this means that the decision-maker should (not) take action if the event is forecast.

The second bracket in Eq. (3) represents the choice between action and inaction, given that the event is *not* forecast to occur. In this case,

for $\alpha_2 > 1-(C/L)$, it follows that $L(1-\alpha_2) < C$, and the rational choice is to not take action. Since α_1 and α_2 are related by Eq. (4), the second threshold corresponds to

$$\alpha_1^{**} = 1 - [(C/L)(1-p)/p] \quad (7)$$

An accuracy below (above) this value means that the decision-maker should (not) take action if the event is not forecast.

It can also be straightforwardly demonstrated that the two critical accuracies are equal when $C/L = p$, while $C/L > p$ corresponds to $\alpha_1^* > \alpha_1^{**}$.

The value of the actual forecast accuracy relative to the critical values determines the choices a rational decision-maker should make. For example, if $\alpha_1 > \max(\alpha_1^*, \alpha_1^{**})$, it follows that the decision-maker should take action when the event is forecast, and not take action when the event is not forecast, giving $EE_{Forecast} = pC + pL[1 - \alpha_1] \leq EE_{CI}$. Thus, $EE_{Forecast}$ is linearly dependent on α_1 for $\alpha_1 > \max(\alpha_1^*, \alpha_1^{**})$. As $\alpha_1 \rightarrow 1$, this expression tends to the limit for perfect information, i.e. $EE_{Forecast} \rightarrow pC = EE_{PI}$.

We can also calculate the values of $EE_{Forecast}$ when the accuracy criterion above is not satisfied. Firstly, $\alpha_1^{**} < \alpha_1 < \alpha_1^*$ implies that $C/L > p$, and it can be shown from Eq. (4) that $EE_{Forecast} = pL = EE_{CI}$, independently of α_1 . Secondly, $\alpha_1^* < \alpha_1 < \alpha_1^{**}$ implies $C/L < p$, and it follows again from equation (4) that $EE_{Forecast} = C = EE_{CI}$ independently of α_1 . Thirdly, for $\alpha_1 < \min(\alpha_1^*, \alpha_1^{**})$ it follows that $EE_{Forecast} = pL\alpha_1 + (1-p)C \leq EE_{CI}$, meaning that $EE_{Forecast}$ is linearly dependent on α_1 for $\alpha_1 < \min(\alpha_1^*, \alpha_1^{**})$. In this limit, $EE_{Forecast} = pC$ if $\alpha_1 = (2p-1)(C/L)$, which is only physically meaningful for $p > 1/2$. Thus, the behaviour of $EE_{Forecast}$ is not symmetrical around $\alpha_1 = C/L$, as would be expected if we had specified that $\alpha_1 = \alpha_2$ (see Appendix).

For the assumptions considered here, the analysis above shows that the critical forecast accuracy is

$$\alpha_{1,crit} = \max(\alpha_1^*, \alpha_1^{**}) \quad (8)$$

2.4. Statistical distribution of Mean Realised Value (MRV)

The Mean Realised Value (MRV) is defined as the mean realised economic value relative to a benchmark, for a forecast system used over a finite time interval, e.g. a seasonal forecast used once annually for 5, 10, 20 or 50 years. If V_n is the realised economic value for the n th year of a sequence, the MRV is written as

$$MRV = (1/N) \sum_{n=1}^N V_n \quad (9)$$

where the realised economic value is defined as

$$V_n = (E_{CI} - E_{Forecast}) / (EE_{CI} - EE_{PI})$$

with E_{CI} and $E_{Forecast}$ being the realised expenses incurred using climatological information and forecast information, respectively.

Due to statistical sampling effects during finite periods of usage, the MRV can differ significantly from the expected economic value in Eq. (1). The statistical distribution of MRV for a given cost/loss ratio, forecast accuracy, climatological frequency and period of usage can, therefore, give important context for user experiences.

Accounting for multiple forecasts per year would be possible within the methodology outlined above, but is beyond the scope of this work. For example, issuing multiple forecasts for a given season could help ensure that the realised performance more closely matches the expected performance. However, the benefit of issuing the additional forecasts will depend on strength of correlations between them which, in turn, depends on the forecast skill. In the trivial limit that the forecasts are perfectly correlated, there is no benefit in issuing more than one per season. In the other extreme, where forecasts are not correlated at all, the forecast system has no skill by definition. This suggests that the greatest benefit of issuing multiple forecasts for a specific season will tend to be for systems with moderate levels of skill.

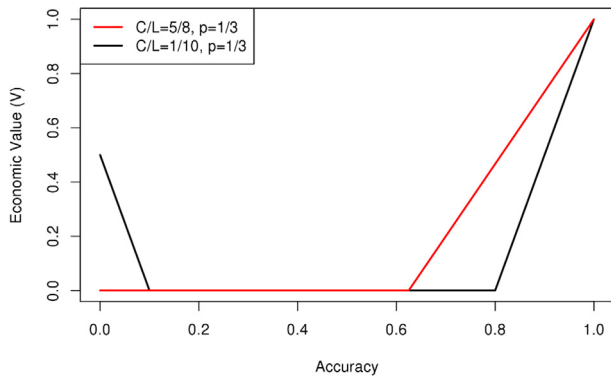


Fig. 1. Economic value (V) of a binary deterministic forecast system, with frequency bias of unity, as a function of accuracy (i.e. α_1) for $C/L = 5/8$ and $1/10$, with $p = 1/3$.

3. Results

3.1. Critical accuracy

Fig. 1 illustrates the expected economic value (V) for the forecast system as a function of accuracy (i.e. α_1), using $C/L = 5/8$ and $1/10$ with $p = 1/3$. Note that the non-zero economic value for low accuracies can only be achieved by the user always doing the opposite of what the forecast says, suggesting that the forecast should not be used in this case.

According to Eqs. (6) and (7), the critical forecast accuracy for $C/L = 5/8$ and $p = 1/3$ is $5/8 = 0.625$ (see red line in **Fig. 1**). Similarly, the critical accuracy for $C/L = 1/10$ and $p = 1/3$ is 0.8 , with a lower critical value of 0.1 (see black line in **Fig. 1**).

The parameter dependence of the critical accuracy is shown in **Fig. 2** – the different functional forms of $\alpha_{1, \text{crit}}^*$ and $\alpha_{1, \text{crit}}^{**}$ are clearly evident, being joined by the $C/L = p$ diagonal. The lack of symmetry around $C/L = p$ is a direct consequence of the frequency bias constraint in Eq. (5). This contrasts with the results if we assume $\alpha_1 = \alpha_2$, for which $\alpha_{1, \text{crit}}$ is symmetrical around the $C/L = p$ diagonal, see Appendix.

In **Fig. 2**, the lowest critical accuracies occur for small cost/loss ratios and where p is no more than marginally greater than C/L . Users and decisions that sit in this part of the parameter space have the best chance of benefitting from a predictive climate service.

In contrast, the critical accuracy is high for both low cost/loss ratios combined with a high climatological probability, and high cost/loss ratios. In this part of the parameter space, it is unlikely that current seasonal forecasting capabilities offer a better alternative than climatological information (or another benchmark). This may change in the

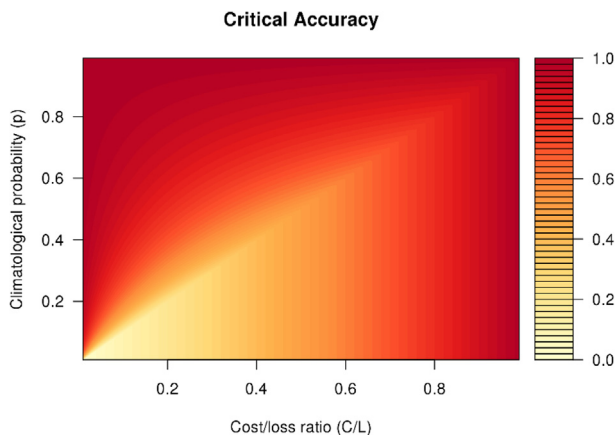


Fig. 2. Critical accuracy ($\alpha_{1, \text{crit}}$), above which EE_{Forecast} is less than EE_{CL} , as a function of the cost/loss ratio (C/L) and climatological probability (p).

future as forecast accuracy continues to improve, but will also depend on the limits of predictability which tend to fall with increasing lead time.

Given the magnitude and range of critical accuracies relative to current capabilities, it is imperative for service developers to work with decision-makers to explore plausible cost/loss ratios in order to assess whether current forecasting capabilities can offer clear benefits to their decision-making.

3.2. Mean realised value

To illustrate the parameter dependence of the MRV distribution, **Figs. 3** and **4** show box and whisker plots for fixed cost/loss ratios ($C/L = 5/8$ and $1/10$) and annual climatological probability ($p = 1/3$), while varying the forecast accuracy from 0.5 to 0.9 and forecast sequence length from 5 to 50 years. These specific cost/loss ratios were chosen to demonstrate implications associated with their magnitude relative to the climatological probability, i.e. $C/L = 5/8 > p$, while $C/L = 1/10 < p$.

The negative skewness evident in the box and whisker plots reflects losses being greater than costs, with the distribution becoming progressively less skewed for longer forecast sequences. This behaviour is broadly expected from the Central Limit Theorem (which implies that the sampled distribution of averages tends to a Gaussian distribution as sample size increases). For example, the probability that a skillful forecast results in “misses” 50 times in a row is much smaller than the probability of 5 “misses” in a row. Similarly, there is a non-zero probability that a user will experience an unbroken sequence of correct rejections – the expenses for this sequence will be 0 , by definition.

The theoretical maximum and minimum values of MRV are shown in **Table 2**, and discount mutually exclusive combinations of realised expenses using both climatological and forecast information. For example, according to the decision-making rules for $C/L > p$, it follows that if a decision-maker using only climatological information incurred a loss, L , then a decision-maker using a skillful forecast for the same event must have incurred either a loss, L , or taken action with cost, C . These limits provide additional context for assessing forecast system performance over a finite period, e.g. indicating the maximum possible losses in advance of using the system.

Evaluating the formulae in **Table 2** for $C/L = 5/8$ with $p = 1/3$, shows that the maximum and minimum values of MRV are 3 and -5 , in agreement with **Fig. 3**. For $C/L = 1/10$ with $p = 1/3$, the maximum and minimum values of MRV are $3/2$ and $-27/2$, in agreement with **Fig. 4**.

The length of a forecast sequence required for the distribution to be approximately Gaussian is a function of the cost/loss ratio (i.e. the initial skewness), with a longer sequence (> 50) needed for a smaller ratio. The width of the distribution also decreases for longer forecast sequences, as the standard error on the mean value reduces.

For accuracies greater than the critical threshold, there is no difference between users of Type 1 and 2. However, for accuracies below the critical threshold the MRV has zero spread for decision-makers of Type 2, as shown in the lower panels of **Figs. 3** and **4**. In this limit, the decisions rely only on climatological information; thus for $C/L < p$, users will always take action with expense, C , while for $C/L > p$, users will always not take action with loss, L or 0 .

There is added complexity for accuracies below the critical threshold combined with $C/L > p$. In particular, Type 2 decision-makers would expect $MRV = 0$ for accuracies below the critical threshold, but could nevertheless experience significant losses or benefits depending on the actual events. In contrast, Type 1 users would expect $MRV < 0$ for accuracies below the critical threshold, as shown in the upper panel of **Fig. 3**. Because of this, risk-averse decision-makers may choose to always take action in order to minimize their maximum losses, particularly if there is little difference in the expected expenses of the two decision-making strategies.

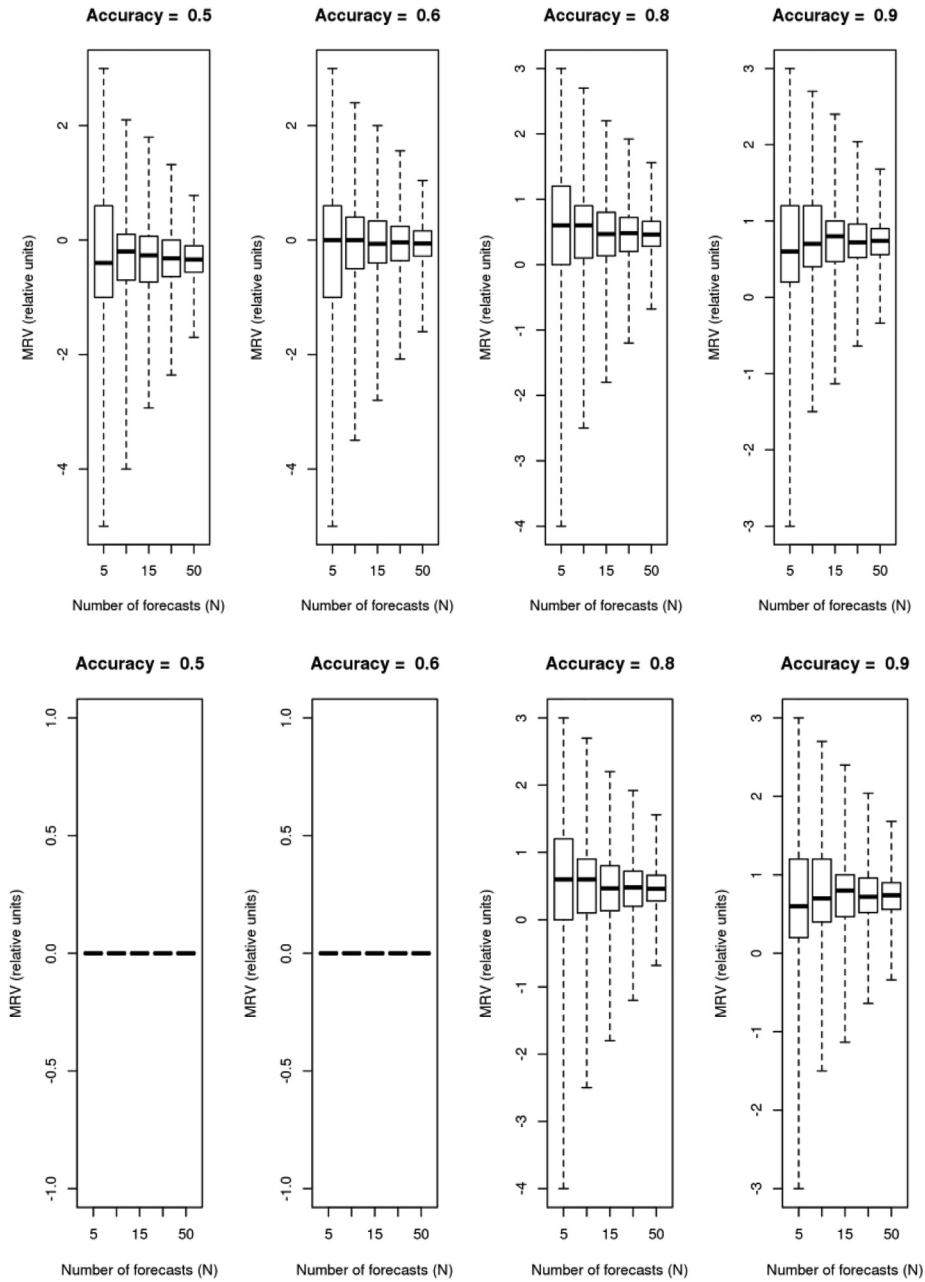


Fig. 3. Box and whisker plots showing the distribution of Mean Realised Value (MRV) as a function of forecast accuracy and number of forecasts (N) for which the system is used. Shown for $C = 5$, $L = 8$, $p = 1/3$, giving $\alpha_{1, \text{crit}} = 0.625$, see Eq. (8). The box shows the 25th, 50th and 75th percentiles, while the whiskers show the full range of the simulated MRV. The theoretical maximum and minimum values of the MRV are 3 and -5 , respectively. Upper panel: decision-maker always follows forecast guidance; lower panel: decision-maker follows forecast guidance when forecast accuracy exceeds critical threshold.

Figs. 5 and 6 show the prior likelihood that $\text{MRV} > 0$, i.e. the prior likelihood that a forecast system used once per year for N years will outperform the climatological benchmark over the same sequence of events. Henceforth, this is referred to as the “exceedance probability”. From a user’s perspective, the exceedance probability should ideally be large (i.e. significantly $> 50\%$) to maximise the chance of obtaining a tangible economic benefit from using the service over a finite period. In each case, we present the exceedance probability for the two types of user described above: i) those who always follow forecast guidance (upper panel), and ii) those who follow guidance when the accuracy exceeds the critical threshold (lower panel).

Even within the context of the deterministic binary forecast considered here, the exceedance probability for users who always follow forecast guidance (upper panel) is not a simple function of forecast

accuracy, the length of the forecast sequence, climatological event probability and the cost/loss ratio. For the combination of parameters in Fig. 5 (i.e. $C/L = 5/8$; $p = 1/3$), the exceedance probability increases both with accuracy and the number of forecasts. Note that the exceedance probability is only substantially > 0.5 for accuracies greater than the critical accuracy (i.e. 0.625 in this example), because the mean of the MRV distribution is only > 0 when the accuracy is greater than the critical value. The contours of constant exceedance probability are shaped by the skewness of the distribution: for accuracies exceeding the critical value, lower negative skewness with increasing number of forecasts leads to a greater exceedance probability.

The lower panel of Fig. 5 demonstrates that, for more sophisticated users, the exceedance probability has a strong cutoff at the critical accuracy of 0.625 .

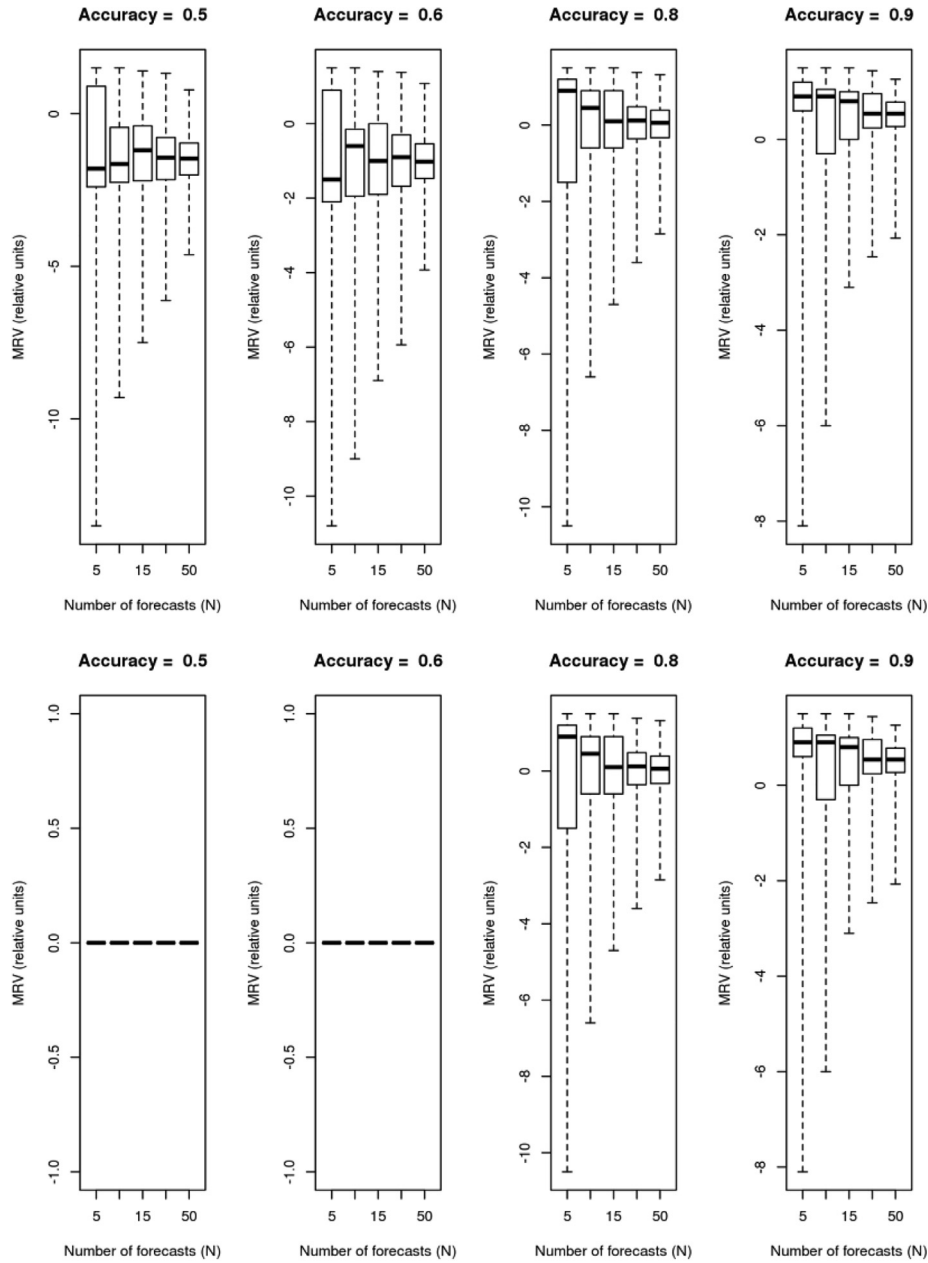


Fig. 4. Box and whisker plots showing the distribution of Mean Realised Value (MRV) as a function of forecast accuracy (α_1) and number of years (N) for which the forecast system is used. Shown for $C = 5$, $L = 50$, $p = 1/3$, giving $\alpha_{1, \text{crit}} = 0.8$, see Eq. (8). The box shows the 25th, 50th and 75th percentiles, while the whiskers show the full range of the simulated MRV. The theoretical maximum and minimum values of the MRV are $3/2$ and $-27/2$, respectively. Upper panel: decision-maker always follows forecast guidance; lower panel: decision-maker follows forecast guidance when forecast accuracy exceeds critical threshold.

Table 2

Theoretical maximum and minimum values of MRV.

	Maximum MRV	Minimum MRV
$C/L > p$	$\frac{1}{p}$	$\frac{1}{[p(1 - (L/C))]}$
$C/L < p$	$\frac{1}{(1-p)}$	$\frac{1}{[1 - (L/C)](1-p)}$

For the combination of parameters shown in Fig. 6, we see a similarly clear disparity in behavior for the two types of decision-maker. For users who always follow forecast guidance (upper panel), the exceedance probability increases with accuracy but is also non-zero for forecast accuracies below the critical value of 0.8. This effect occurs because the wide spread in MRV at all accuracies ensures there is a non-

zero probability of a low-skill forecast system outperforming climatological information over a finite period.

Like Fig. 5, the lower panel of Fig. 6 exhibits a cutoff at the critical accuracy value (0.8 in this case), associated with the transition between using climatological information and skillful forecast information.

The rippling effect evident in Fig. 6 is a feature of the cost/loss model and relates to the sampling rates of the four different decision-outcome combinations over a finite sequence of forecasts. This real effect translates into changes in the skewness of the MRV distribution for different numbers of forecasts. In turn, the skewness of the distribution modulates the probability that $\text{MRV} > 0$, leading to the ripples.

To see this, it is helpful to consider the main factors that affect the number of occurrences of a decision-outcome combination as a function of forecast sequence length. Assume that the annual probability of

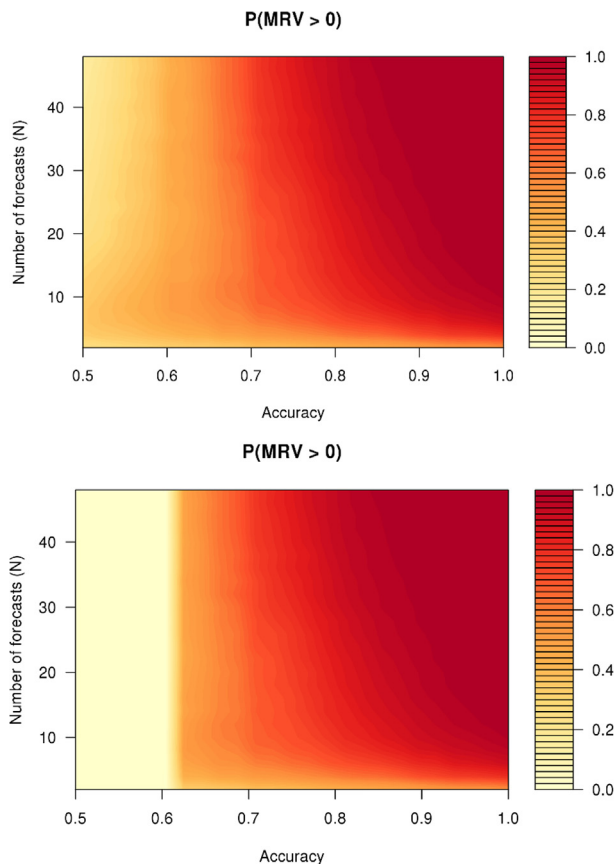


Fig. 5. Prior likelihood that the Mean Realised Value (MRV) will exceed 0, as a function of forecast accuracy (α_1) and number of forecasts (N) for which the forecast system is used. Shown for $C = 5$, $L = 8$, $p = 1/3$, giving $\alpha_{1, \text{crit}} = 0.625$, see Eq. (8). Upper panel: Type 1 decision-makers who always follow forecast guidance; lower panel: Type 2 decision-makers who follow forecast guidance when forecast accuracy exceeds critical threshold.

obtaining one of the four decision-outcome combinations is w . Then, for a forecast sequence of length N , the expected number of occurrences for this decision-outcome combination is a continuous quantity, Nw . However, the median number of times this combination occurs in 10,000 forecast sequences of length N , must always be an integer. If $N < 1/(2w)$, the median number of occurrences is exactly 0, since it is more likely that this decision-outcome combination will not occur. Similarly, if $1/(2w) < N < 3/(2w)$, the median number of occurrences is exactly 1. In this example, there are discontinuities at $N = 1/(2w)$ and $N = 3/(2w)$, where the median switches first from floor(Nw) to ceil(Nw), and then back.

The difference between the mean and median is a useful proxy for skewness. Consequently, for $q < Nw < q + 0.5$, it follows that the mean $>$ median, and that the skewness is positive; similarly, for $q + 0.5 < Nw < q + 1$, it follows that the mean $<$ median, and the skewness is negative, where q is an integer. The overall skewness for the MRV distribution is determined by this effect applied across all four decision-outcome combinations in the 10,000 forecast sequences, incorporating the costs and losses for each combination.

For short forecast sequences, the skewness of the MRV distribution is dominated by the cost/loss ratio. The sampling effect for different decision-outcome combinations is superimposed on this and slightly increases or decreases the skewness. For a forecast accuracy less than the critical threshold, the probability that a random forecast sequence outperforms the expectation value is lower for a positively skewed distribution than for a negatively skewed distribution, assuming a constant mean.

The sampling effect is comparatively large for $N \sim 1/w$, and

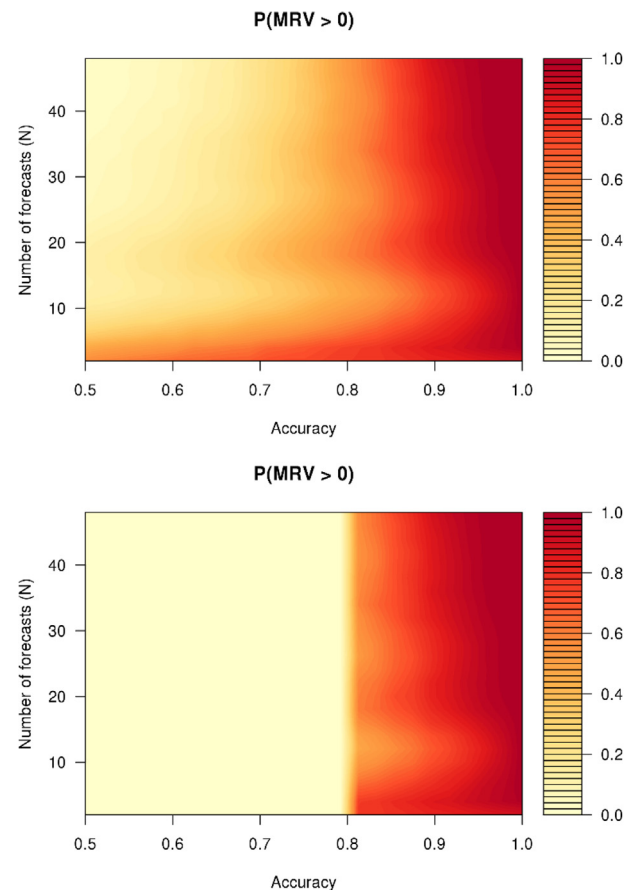


Fig. 6. Prior likelihood that the Mean Realised Value (MRV) will be > 0 , as a function of forecast accuracy and number of forecasts (N) for which the forecast system is used. Shown for $C = 5$, $L = 50$, $p = 1/3$, giving $\alpha_{1, \text{crit}} = 0.8$, see Eq. (8). Upper panel: Type 1 decision-makers who always follow forecast guidance; lower panel: Type 2 decision-makers who follow forecast guidance when forecast accuracy exceeds critical threshold.

becomes negligible in the limit that $N \gg 1/w$, since the incremental changes in the median number of occurrences become smaller in relative terms. In general, a forecast miss (with an accompanying loss, L), will be the rarest category, meaning that the associated ripples are likely to be evident for relatively large numbers of forecasts, depending on the forecast accuracy. While this effect is a feature of the idealized cost/loss model, it is unclear whether it should be a consideration in real decision-making, where decisions may not be discrete.

These examples highlight some of the complexities in assessing user-relevant performance metrics, even for a deterministic binary forecast with fixed system parameters and rational users. Nevertheless, the results indicate the importance of exploring the decision-dependent critical accuracy and the distribution of realised economic value rather than its expectation.

More generally, it is clear that, for any finite sequence of skillful but imperfect forecasts, there is a non-zero probability that the realised expenses will exceed those based on using only climatological information. This suggests that risk-averse decision-makers acting over a finite period may prefer to take mitigating action that minimises their maximum losses, largely regardless of whether $C/L < p$ or not.

4. Summary

The increasing interest in climate services, and their potential for widespread use, highlights the need to ascertain their potential benefits and limitations. In particular, it is necessary to assess the performance of climate services against criteria that directly reflect utility to the

user, not just the standard metrics for weather/climate models (e.g. [Golding et al, 2017](#)). Doing so is an essential component of ensuring that climate services are fit for purpose, and building user's confidence in the underpinning weather/climate science and its application to real world challenges (e.g. [Mason, 2004](#); [Hewitt et al, 2018](#); [Vaughan et al, 2018](#)).

Within this context, we have explored two important measures of user-relevant performance through combining a decision theoretic approach with stochastic modelling of a deterministic binary forecast (for which the frequency bias is unity). These are:

- 1) The parameter dependence of the critical forecast accuracy above which an imperfect forecast-based service provides economic benefit relative to a benchmark - represented, in this case, by climatological information.
- 2) The probability distribution of "Mean Realised Value" (MRV) over a finite sequence of forecasts (e.g. issued once annually for 5, 10, 15, 25 and 50 years), and as a function of the cost/loss ratio, climatological event probability and forecast accuracy. This highlights the non-negligible consequences of statistical sampling effects which mean that, over a short sequence of forecasts, the realised economic value can differ significantly from the expectation value. Through simulating the distribution of MRV, we can evaluate the prior likelihood that a climate service will outperform decision-making based on climatological information.

Regarding the first point, the critical accuracy is shown to be a function of only the cost/loss ratio and the climatological frequency of the hazard. In this case, the critical accuracy is largest for high cost/loss ratios (i.e. costs only slightly less than losses), and when the climatological probability exceeds the cost/loss ratio (see [Fig. 2](#)). Comparing current forecasting capabilities with estimates of a decision-dependent critical accuracy may be a convenient way of exploring the potential usefulness of a climate service to a user. Specifically, forecasts which are more (less) accurate than a critical value, do (not) on average provide information which is more useful than the benchmark.

However, as highlighted by the second point above, the realised economic value can differ considerably from the expectation over finite periods of usage, which can affect user interpretation of forecast capabilities. By stochastically simulating 10,000 N-year forecast sequences we have generated the statistical distribution of MRV as a function of the cost/loss ratio, climatological frequency, forecast accuracy and the period of usage. In turn, this distribution gives the prior likelihood that the service, over a limited time interval, will outperform a user-defined benchmark.

The results presented here vary according to whether decision-makers always follow the forecast guidance, or whether they only take mitigating action when the forecast accuracy exceeds the critical value for their cost/loss ratio. Specifically, users are more likely to experience

significant losses if they make decisions based on forecasts for which the accuracy is below the critical threshold.

More generally, for smaller cost/loss ratios (i.e. losses far exceeding costs), the distribution of MRV is more negatively skewed, meaning that a large loss is much more likely than a large benefit over a finite period of usage. For both types of user, as the forecast system is used for longer, the MRV distribution becomes more Gaussian, as expected from the Central Limit Theorem, while the width of the distribution decreases according to the definition of the standard error on the mean. As demonstrated in [Figs. 5 and 6](#), the prior likelihood that the $MRV > 0$ increases with accuracy, and can depend on the skewness of the distribution (primarily affected by the cost/loss ratio, duration of usage and decision-making approach), and the forecast accuracy relative to the critical value (determined by the climatological probability and cost/loss ratio). These results show that, for any finite sequence of skillful but imperfect forecasts, there is a non-zero probability that the realised expenses will exceed those based on using only climatological information. This suggests that risk-averse decision-makers acting over a finite period may prefer to take mitigating action that minimises their maximum losses, almost irrespective of the cost/loss ratio and climatological probability.

While these results have been generated by simulating a deterministic, binary forecast, with two different approaches for implementing forecast guidance, the general concepts and approach apply more broadly. As such, this approach can help users assess the potential risks associated with adopting a climate service for a given period, and may help identify geographical regions and meteorological parameters for which a forecasting tool could provide a worthwhile investment with respect to the next-best alternative. Doing so will help develop more robust climate services, and maximise the societal benefits of skillful weather forecasts and climate predictions.

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by the Met Office Hadley Centre Climate Programme (GA01101).

Whilst at times it can be difficult to attribute the contribution of each author to a scientific paper, in this case we feel it is important to highlight the leading role Edward Pope had in the formulation of the original idea and its implementation. The contributions of the other two authors have been limited to critical advice comments and discussions which, despite their relative importance in the formulation of the final manuscript, cannot be compared to the work done by the first author.

The authors would like to thank the anonymous reviewers for comments and suggestions that significantly improved this work.

Appendix. – Alternative model assuming equal accuracies for binary forecast categories

The results in this paper are derived under the assumption that the forecast frequency bias is unity, which imposes a relationship between $Pr(Y_1|X_1)$ and $Pr(Y_2|X_2)$ that also depends on the climatological probability, p . For completeness, we demonstrate here the corresponding results under the alternative assumption that $Pr(Y_1|X_1) = Pr(Y_2|X_2) = \alpha$, and that the frequency bias is not unity.

Under this assumption, $Pr(X_k) \neq Pr(Y_k)$, and we assume that the forecast rate is linked to the climatological probability (base rate) by $Pr(X_i) = fPr(Y_i) = fp$, where f is the frequency bias. As before, the definition of the joint probabilities requires $\sum_{k=1}^2 Pr(Y_j \& X_k) = \sum_{k=1}^2 Pr(Y_j|X_k)Pr(X_k) = Pr(Y_j)$; thus, for $j = 1$ we have $Pr(Y_1) = p = fp\alpha + (1 - fp)(1 - \alpha)$. Rearranging this expression shows that the frequency bias and forecast accuracy are related by

$$f = [p + \alpha - 1]/[p(2\alpha - 1)] \quad (A1)$$

As such, the frequency bias is only greater than zero for $\alpha > \max(1 - p, \frac{1}{2})$, or $\alpha < \min(1 - p, \frac{1}{2})$ and is undefined at $\alpha = 1/2$. Taking $p = 1/3$, we find that this model is *not* physically meaningful for $\frac{1}{2} \leq \alpha \leq 2/3$. The restricted parameter space for which this derivation is valid highlights the benefits of assuming the frequency bias is unity, as in the main paper.

In the case considered here $Pr(Y_1|X_1) = Pr(Y_2|X_2) = \alpha$, and equation (4) becomes

$$EE_{Forecast} = fpmin\{C, L\alpha\} + (1 - fp)min\{C, L(1 - \alpha)\} \quad (A2)$$

Following the same arguments as in the main paper, it is evident that there are critical values of the forecast accuracy at $\alpha^* = C/L$ and $\alpha^{**} = 1 - \frac{C}{L}$. Note that these critical thresholds are the same as those for α_1 and α_2 in the main paper; however, in that case the thresholds were derived under the assumption that $\alpha_1 \neq \alpha_2$. The critical thresholds for equation A(2) are symmetrical around $C/L = 1/2$, because the minimum forecast accuracy for this assumption is $\alpha = 1/2$ by construction.

If $\alpha > \max(\alpha^*, \alpha^{**})$, it follows from Eq. A(2) that $EE_{Forecast} = fpC + (1 - fp)L[1 - \alpha] \leq EE_{CI}$. Thus, $EE_{Forecast}$ is dependent on α for $\alpha > \max(\alpha^*, \alpha^{**})$, and tends to the limit for perfect information as $\alpha \rightarrow 1$, i.e. $EE_{Forecast} \rightarrow pC = EE_{PI}$. The dependence is not perfectly linear since the frequency bias is a function of forecast accuracy.

For $\alpha^* < \alpha < \alpha^{**}$, it can be shown from Eq. A(2) that $EE_{Forecast} = pL = EE_{CI}$, independently of α . Similarly, if $\alpha^* < \alpha < \alpha^{**}$ it follows again from Eq. A(2) that $EE_{Forecast} = C = EE_{CI}$ independently of α .

Finally, for $\alpha < \min(\alpha^*, \alpha^{**})$, Eq. A(2) leads to $EE_{Forecast} = fpL\alpha + (1 - fp)C \leq EE_{CI}$, meaning that $EE_{Forecast}$ is dependent on α for $\alpha < \min(\alpha^*, \alpha^{**})$. In this limit, $EE_{Forecast} \rightarrow pC$ as $\alpha \rightarrow 0$ and it can be seen that the behaviour of $EE_{Forecast}$ is symmetrical around $\alpha = 1/2$.

References

- Bett, P.E., Scaife, A.A., Li, C., Hewitt, C., Golding, N., Zhang, P., Dunstone, N., Smith, D.M., Thornton, H.E., Lu, R., Ren, H., 2017. Seasonal forecasts of the summer 2016 Yangtze River basin rainfall. *Adv. Atmos. Sci.*
- Blench, R., 1999. Seasonal climate forecasting: Who can use it and how it should be disseminated? *Nat. Resour. Perspect.* 47, Overseas Development Institute, London. Available at.
- Brooks, M.S., 2013. Accelerating innovation in climate services: the 3 E's for climate service providers. *Bull. Am. Meteor. Soc.* 94, 807–819. <https://doi.org/10.1175/BAMS-D-12-00087.1>.
- Bruno Soares, M., Dessai, S., 2015. Exploring the use of seasonal climate forecasts in Europe through expert elicitation. *Clim. Risk. Manag.* 10, 8–16.
- Buontempo, C., Hanlon, H., Bruno Soares, M., Christel, I., Soubeyroux, J.-P., Viel, C., Calmanti, S., Bosi, L., Falloon, P., Palin, E., Vanvyve, E., Torralba, V., Gonzalez-Reviriego, N., Doblas-Reyes, F., Pope, E., Liggins, F., Newton, P., 2017. What have we learnt from EUPORIAS climate service prototypes? *Clim. Services*. <https://doi.org/10.1016/j.cliser.2017.06.003>.
- Falloon, P., Soares, M.B., Manzanar, R., San-Martin, D., Liggins, F., Taylor, I., Kahana, R., Wilding, J., Jones, C., Comer, R., de Vreede, E., Som de Cerff, W., Buontempo, C., Brookshaw, A., Stanley, S., Middleham, R., Pittams, D., Lawrence, E., Bate, E., Peter, H., Uzell, K., Richards, M., 2018. The land management tool: developing a climate service in Southwest UK. *Clim. Serv.* 9, 86–100.
- Goddard, L., et al., 2010. Providing seasonal-to-interannual climate information for risk management and decision-making. *Procedia Environ. Sci.* 1 (2010), 81–101.
- Golding, N., Hewitt, C., Zhang, P., Bett, P., Fang, X., Hu, H., Nobert, S., 2017. Improving user engagement and uptake of climate services in China. *Clim. Serv.* 5, 39–45. <https://doi.org/10.1016/j.cliser.2017.03.004>.
- Henley, E.M., Pope, E.C.D., 2017. Cost-loss analysis of ensemble solar wind forecasting: space weather use of terrestrial weather tools. *Space Weather* 15. <https://doi.org/10.1002/2017SW001758>.
- Hewitt, C., Mason, S., Walland, D., 2012. The global framework for climate services. *Nat. Clim. Change* 2 (11), 831–832.
- Hewitt, C., Stone, R., Tait, A., 2017. Improving the use of climate information in decision-making. *Nat. Clim. Change* 7 (2017), 614–616. <https://doi.org/10.1038/nclimate3378>.
- Katz, R., Murphy, A., 1997. Forecast value: prototype decision making models. In: Katz, R., Murphy, A. (Eds.), *Economic value of weather and climate forecasts*. Cambridge.
- Lindley, D.V., 1985. *Making decisions*, 2nd edn. Wiley, New York, NY.
- Mason, I., 2004. The cost of uncertainty in weather predictions: modeling quality-value relationship for yes/no forecasts. *Aust. Meteorol. Mag.* 53, 111–122.
- Meza, F.J., Hansen, J.W., Osgood, D., 2007. Economic value of seasonal climate forecasts for agriculture: review of ex-ante assessments and recommendations for future research. *J. Appl. Meteorol. and Climatology* 47.
- Mylne, Kenneth R., 2000. Decision-making from probability forecasts based on forecast value. *Meteorol. Appl.* 9 (3), 307–315. <https://doi.org/10.1017/S1350482702003043>.
- Palin, E.J., Scaife, A.A., Wallace, E., Pope, E.C.D., Arribas, A., Brookshaw, A., 2016. Skillful seasonal forecasts of winter disruption to the UK transport system. *J. Appl. Meteorol. Climatol.* 55, 325–344.
- Palmer, T.N., 2002. The economic value of ensemble forecasts as a tool for risk assessment: from days to decades. *Q.J.R. Meteorol. Soc.* 128, 747–774.
- Pope, E.C.D., Buontempo, C., Economou, T., 2017. Quantifying how user-interaction can modify the perception of the value of climate information: a Bayesian approach. *Clim. Serv.* 6, 41–47. <https://doi.org/10.1016/j.cliser.2017.06.006>.
- Richardson, D., 2000. Skill and relative economic value of the ECMWF Ensemble Prediction System. *Q. J. Roy. Meteor. Soc.* 126, 649–668.
- Richardson, David S., 2011. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, Ltd, Chichester, UK, pp. 167–184. <https://doi.org/10.1002/9781119960003.ch9>.
- Thornes, J.E., Stephenson, D.B., 2001. How to judge the quality and value of weather forecast products. *Meteorol. Appl.* 8, 307–314.
- Suckling, E., van Oldenborgh, G., Eden, J., Hawkins, E., 2016. An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts. *Clim. Dyn.* 48, 3115–3138.
- Van den Dool, H., 2006. *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press.
- Vaughan, C., Dessai, S., Hewitt, C., 2018. Surveying climate services: what can we learn from a bird's-eye view? *Weather, climate and society* 10, 373–395. <https://doi.org/10.1175/WCAS-D-17-0030.1>.
- Wilks, D.S., 2001. A skill score based on economic value for probability forecasts. *Meteorol. Appl.* 8, 209–219.